View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Institutional Knowledge at Singapore Management University

Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of Business

Lee Kong Chian School of Business

12-2012

Designing selection systems for medicine: The importance of balancing predictive and political validity in high-stakes selection contexts

Fiona PATTERSON
University of Cambridge

Filip LIEVENS
Singapore Management University, filiplievens@smu.edu.sg

Máire KERRIN

Lara ZIBARRAS
City University

Bernd CARETTE
Ghent University
DOI: https://doi.org/10.1111/ijsa.12011

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the Human Resources Management Commons, and the Organizational Behavior and Theory Commons

Citation

PATTERSON, Fiona; LIEVENS, Filip; KERRIN, Máire; ZIBARRAS, Lara; and CARETTE, Bernd. Designing selection systems for medicine: The importance of balancing predictive and political validity in high-stakes selection contexts. (2012). *International Journal of Selection and Assessment*. 20, (4), 486-496. Research Collection Lee Kong Chian School Of Business.
Available at: https://ink.library.smu.edu.sg/lkcsb_research/5641

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Published in International Journal of Selection and Assessment, 2012 December, Volume 20, Issue 4, Pages 486-496
http://doi.org/10.1111/ijsa.12011

Designing Selection Systems for Medicine: The Importance of Balancing Predictive and Political Validity in High Stakes Selection Contexts

Professor Fiona Patterson, University of Cambridge, UK

Professor Filip Lievens, University of Ghent, Belgium.

Dr Máire Kerrin, Work Psychology Group Ltd, UK.

Dr Lara Zibarras, City University, London, UK.

Dr. Bernd Carette, University of Ghent, Belgium.

In press *International Journal of Selection and Assessment*

Corresponding author:

Professor Fiona Patterson, email: f.patterson@workpsychologygroup.com

Department of Social and Developmental Psychology

University of Cambridge

Free School Lane

Cambridge

CB2 3RQ

Abstract

Selection into medical education and training is a high stakes process. A key unanswered issue is the effectiveness of measuring noncognitive predictors via both low-fidelity and high-fidelity selection approaches in this high-stakes context. We review studies investigating the effectiveness of multiple selection instruments in terms of predictive validity, incremental validity, and applicant reactions in both entry-level and advanced level medical selection. Our results show that the SJT is the best single predictor of performance, operationalized in multiple ways. In addition, the low fidelity SJT has incremental predictive power over cognitively oriented tests, and high fidelity AC exercises add incremental validity over the low fidelity (and less costly) selection methods. Concerning applicant reactions, results show that overall the selection system is positively received. However, the method with the highest predictive validity – the SJT – received comparatively lower face validity ratings which may present a "*justice dilemma*" for employers. Furthermore, various other stakeholders have a political interest in the selection methods used (eg government, the regulators and trade unions).

Designing Selection Systems for Medicine: The Importance of Balancing Predictive and Political Validity in High Stakes Selection Contexts

Medical selection is high stakes, both in terms of the human and financial costs. For students, being selected to medical school might determine their future career and life. Similarly, the advanced-level medical selection process (e.g., selection into medical specialties such as General Practice) is high-stakes because physicians have already completed many years of training and therefore have already invested substantial time and effort in their career planning. Internationally, medical selection practices are also characterized by high volume and high competition. In the UK, for example, there are over 25,000 applicants for 8,000 medical schools places each year and over 12,000 applicants for 8,000 specialty training posts. Due to its high-stakes, high-volume nature and impact on society, medical selection is also a high profile process that attracts significant public interest and media attention (Irish, Carr, Sowden, Douglas, & Patterson, 2011).

Generally, there are two main selection gateways within the medical training pathway, namely selection into medical school (i.e., *entry-level* selection, pre-employment selection) and selection into postgraduate training (i.e., *advanced level* selection, employment selection). Historically, these two major medical selection gateways have been relatively untouched by research on the predictive validity of selection methods. For example, in some countries, selection systems for medicine typically relied on unstructured panel interviews and multiple referee reports (Jefferis, 2007). In other countries, medical school admissions systems were based on lottery systems (Patterson & Ferguson, 2007). Furthermore, selection practices in medicine were often characterized as encouraging nepotism and patronage and influenced by current socio-political agendas (Greenhalgh, 2010).

Over the last years, this state-of-affairs in medical selection has changed as internationally there have been several examples wherein medical selection has benefited from

insights in the broader field of personnel selection or has been enabled to experiment with new innovative selection approaches. The majority of this research has focused on entry level selection (e.g., McManus, Smithers, Partridge, Keeling, & Fleming, 2003; McManus, Powis, Wakeford, Ferguson, James, & Richards, 2005) but there is now an emerging literature on advance level selection (see Prideaux et al., 2011). So, the description of medical selection being a relatively under-researched occupational setting is no longer accurate. One such a striking example has been the attempt to measure a broader range of competencies in medical selection beyond academic abilities. Typically, this attempt involves measuring both cognitive (e.g. clinical knowledge base) and noncognitive (e.g. empathy, integrity) competencies in medical selection. To measure noncognitive competencies, both more expensive *high-fidelity* methods (e.g., assessment center, AC, exercises, which directly replicate job tasks and require an actual behavioral response from applicants) and relatively inexpensive *low-fidelity* selection methods (e.g., situational judgment tests, SJTs, which elicit multiple choice responses to written or video-based situations, Motowidlo, Dunnette, & Carter, 1990) have been proposed as viable formal measurement strategies.

A key unanswered question is whether these low-fidelity and/or high-fidelity selection approaches indeed "deliver" what they attempt to do in both entry-level and advanced-level medical selection contexts, namely adding incremental predictive power over and above cognitively-oriented selection predictors. Similarly, it is important to compare the predictive validity of the inexpensive low-fidelity selection approaches to that one of the more expensive high fidelity simulation approaches. Equally, in the high profile medical selection context we need to find out how applicants react to these various noncognitive selection instruments and whether their reactions are more favorable as compared to the traditional cognitively oriented tests.

The aim of this paper is to provide evidence-based answers to these crucial questions by reviewing recent research about the effectiveness of measuring noncognitive predictors via low-fidelity (SJT) and/or high-fidelity (AC) selection methods in both entry-level and advanced level medical selection medical selection. In our review, "effectiveness" will be broadly defined. That is, we examine the low-fidelity and high-fidelity selection methods in terms of traditional criteria such as predictive validity and incremental validity. In addition, we examine candidate perceptions and explore other stakeholder reactions towards both the low-fidelity and high-fidelity selection methods as an important criterion to consider (see Anderson, Salgado, & Hulsheger, 2010) in light of the high-stakes context of medical selection practices.

Our review focuses on medical selection research about the effectiveness of low-fidelity (SJT) and/or high-fidelity (AC) selection methods in primarily two countries, namely the UK and Belgium. We decided to concentrate on these two countries because they enable us to examine the effectiveness of introducing noncognitive predictors that "go beyond" traditional cognitive tests in both an entry-level (Belgium) and advanced level (UK) high-stakes context. In particular, in the UK, both high-fidelity (i.e., assessment center exercises) and low-fidelity (SJTs) selection methods have been introduced in addition to more cognitively-oriented selection predictors for selecting General Practitioners. Conversely, to our knowledge, Belgium is the only country where low-fidelity selection methods (SJTs) have actually been used in addition to a cognitively-oriented test battery in a medical school admissions (entry level) context. In other countries (e.g., the USA) SJTs have been examined only in a low stakes educational context (i.e., for research purposes, see Oswald, Schmitt, Kim, Ramsay, & Gillespie, 2004; Schmitt et al., 2009).

The research evidence as well as its implications reviewed in this paper are important not only for medical selection, but also have the potential to inform the field of personnel selection in general. First, the trend of investing in predictors that "go beyond" cognitively-oriented

selection predictors has also been notable in other educational selection domains and in the employment area (Sackett & Lievens, 2008). Second, in personnel selection in general there is a scarcity of research evidence on the comparative validity of low-fidelity assessments (e.g., SJTs) versus high-fidelity assessments (e.g., AC exercises). Third, the high-stakes nature of medical selection puts other contextual factors (e.g., test coaching, public scrutiny of the selection process) on the research agenda that are under-researched in the broad field of personnel selection.

As the backdrop of the above, our review is structured around the following questions:

1. *What is the predictive validity of low-fidelity and high-fidelity simulations in advanced level selection?*
2. *What is the predictive validity of low-fidelity simulations in entry-level selection?*
3. *What is the incremental validity of low-fidelity and high-fidelity simulations over cognitively-oriented predictors in advanced level selection?*
4. *What is the incremental validity of low-fidelity simulations over cognitively-oriented predictors in entry-level selection?*
5. *How do applicants react towards low-fidelity and high-fidelity selection instruments in high-stakes settings?*

Question 1: What Is The Predictive Validity Of Low-Fidelity And High-Fidelity Simulations In Advanced Level Selection?

Several meta-analyses have shown that SJTs are effective measures of future job performance (Christian, Edwards, & Bradley, 2010; McDaniel, Hartman, Whetzel, & Grubb, 2007; McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001), with corrected validities ranging from .26 to .34. Similarly, ACs have a good psychometric track record in terms of predictive validity (Arthur, Day, McNelly, & Edens, 2003). However, the majority

of these studies, especially for SJTs, were conducted in concurrent settings (Whetzel & McDaniel, 2009). Due to the high competition and high stakes in medical selection, a unique set of issues arise which require investigation into the generalizability of the validities found in current studies to actual selection settings. For example, in high-stakes contexts, job applicants may be more motivated to distort their responses (Whetzel & McDaniel, 2009). Furthermore, in many countries there exists an industry of independent coaching firms whose sole purpose is to help applicants be successful in the selection process. It can also be assumed that the content of a test will get known[1] and be distributed among candidates.

To date, the predictive validity of low-fidelity and high-fidelity simulations in an advanced level medical selection context has primarily been investigated in the UK. The advanced level selection methodology in the UK entails three selection stages (see Figure 1). Following stage one eligibility checks, criteria are assessed at a shortlisting stage two via two bespoke selection tests: a clinical problem solving test, where candidates apply clinical knowledge to solve a problem reflecting a diagnostic process or developing a management strategy for a patient; and a situational judgment test, where candidates are presented with written work-related scenarios relating to professional dilemmas to which they have to choose an appropriate response from a list of alternatives (Patterson, Baron, Carr, Plint, & Lane, 2009; Patterson, Carr, et al., 2009). The SJT items are designed to target several non-cognitive attributes including empathy, integrity and coping with pressure. Table 1 provides sample items of both the SJT and the clinical problem solving test. Subsequently, stage three is an assessment center that includes three selection methods: a group exercise which involved a group discussion exercise relating to a work-related issue; a simulated patient consultation where candidates play

[1] There have been major incidents of legal case initiation due to security breaches of selection tests for medical schools admissions. For example, in 2009, the Brazilian national entrance exam was stolen - there was a public outcry when two men working for the company hired by the Education Ministry to print the test, stole it, and tried to sell the questions a week before the exam was due to take place. The men were arrested after a newspaper alerted police, and the Education Ministry was forced to redo the entire exam, causing it to be delayed by several weeks.

the role of physician and an actor plays the patient role; and a written in-tray exercise where candidates prioritize a list of work-related issues and justify their choices (Patterson, Baron, et al., 2009; Patterson, Ferguson, Norfolk, & Lane, 2005). If the candidates successfully passed these three stages, they can enter the training program.

Patterson, Baron et al., (2009) conducted a first study wherein they provided validity evidence of the clinical problem solving test and the SJT in predicting assessment center performance, operationalized as the total score across the three assessment center exercises. It should be noted that the findings reported by Patterson, et al (2009) used pilot data from 2006, whilst the data reported here used data from the full cohort in 2007. A significant correlation was found between both tests (clinical problem solving test and the SJT, $r = .53$). Assessment center performance correlated positively with both the clinical problem solving test ($r = .41$) and the SJT ($r = .49$). In line with the findings reported by Patterson, Baron et al. (2009), the SJT was the best predictor of AC performance. In a follow-up study, Koczwara et al. (2012) recently replicated this finding concerning the superiority of the SJT ($r = .50$ - $.54$) in predicting AC performance.

As these studies used only an internal criterion (i.e., assessment center performance), Lievens and Patterson (2011) focused on the validity of the full range of advanced level selection methods (i.e., clinical problem solving test, SJT, assessment center) in predicting an external criterion, namely job performance. The validity evidence of the clinical problem solving test, the SJT, and the assessment center in predicting job performance after one year was analyzed, where the outcome variable was measured using supervisor ratings of trainee performance on a number of performance dimensions (such as clinical expertise, empathy, communication). Their results showed that all of the three selection procedures, problem solving test ($r = .54$), SJT ($r = .56$), and AC ($r = .50$), emerged as significant predictors of job performance.

A final yet unpublished study (Patterson, 2011) examined whether the clinical knowledge test and the SJT could predict end of training competence. After three years of training, general practitioners must complete a final membership examination (similar to a certification exam) so that they can practice independently. This examination includes two parts: (1) an applied knowledge test which includes questions regarding clinical medicine, critical appraisal and evidence-based practice, and health informatics and (2) a clinical skills assessment, which entails a dozen high fidelity patient simulations including professional, clinical, communication and practical skills appropriate for a general practitioner.

Table 2 presents correlations between the predictors and the criterion measures. Correlations corrected for range restriction are also presented. All selection methods are significant predictors for both examinations (corrected *r* ranges from .41 to .85), showing that each of the methods have independent predictive validity (see Table 2). There is a particularly strong correlation between the clinical problem solving test administered when entering the advanced level program and the applied knowledge test administered three years later ($r = .85$), reflecting the similarity between the two selection methods, both being tests of knowledge. Whilst the SJT correlated to a similar extent with both the applied knowledge test ($r = .69$) and the clinical skills assessment ($r = .57$), the clinical problem solving test had a much smaller correlation with the clinical skills assessment ($r = .55$) than with the applied knowledge test.

Question 2: What Is The Predictive Validity Of Low-Fidelity Simulations In Entry-Level Selection?

Whereas the prior research question focused on the validity of low-fidelity (SJTs) and high-fidelity (assessment center) simulations in advanced-level medical selection, one might also wonder whether it is not possible to use these simulations and especially the less expensive low-fidelity simulations in entry-level medical selection (i.e., admission of

students to medical school) where applicants have not acquired any medical (job-specific) knowledge. As SJTs present applicants with depictions of hypothetical scenarios (in either written or video-based format) and ask them to identify a response from a list of alternatives they can be administered to large groups (Lievens, Peeters, & Schollaert, 2008).

Generally, there are two rationales behind the use of SJTs in an entry-level medical (admissions) context. First, even though students at the time of admission might not have any experience with physician-patient situations from the perspective of the physician, it is assumed that their answers on basic interpersonal situations with patients might serve as precursors of their behavior in future actual interactions with patients, as observed/rated during internships and in the job many years later. Second, given that SJTs measure students' initial procedural knowledge about interpersonal behavior at the time of admission, it is assumed that subsequent training on procedural knowledge and skills during medical education might build on that, thereby underscoring (instead of negating) the value of selecting students in this area at the outset.

Empirical evidence from the US shows that SJTs significantly correlated with first-year GPA (r = .16), college absenteeism (r = -.27), self-ratings of college performance (r = .53), and peer ratings of college performance (r = .16). Recently, Schmitt et al. (2009) extended these findings by showing that SJTs correlated .21 with GPA measured four years later in time. However, both of these studies were conducted in a research context. Lievens, Buyse, and Sackett (2005a) investigated the validity of the SJT for actual entry-level medical selection in Belgium. In that country, candidates have to pass a centralized admission exam to gain access to medical education (so there is no selection on the part of the universities). There are two main parts. The cognitive part is a combination of four science knowledge tests (biology, chemistry, mathematics, and physics) and a general mental ability test. Another part assesses interpersonal skills via a video-based SJT depicting 30 physician-patient scenarios.

These scenarios relate to the "building and maintaining relationships" and "exchanging information" components of interpersonal skills (Klein, DeRouin, & Salas, 2006) and were developed with the help of experienced physicians. No medical knowledge is necessary to complete the SJT items. Their results show that scores on the video-based SJT at the time of admission predicted GPA on interpersonal skills courses throughout the curriculum, with validity coefficients ranging between .12 and .55. Conversely, the video-based SJT was neither a significant predictor of first-year GPA nor of medical course grades. For those criteria, the cognitive part emerged as the best predictor. For instance, the correlation between this cognitive part and first-year GPA equaled .52.

Recently, a follow-up study tracked students from admission to employment (Lievens & Sackett, 2012). It was found that the SJT also predicted internship performance seven years later and job performance as a trainee physician nine years later. In particular, validities of the interpersonal SJT for predicting overall internship performance and supervisory-rated job performance were .22 and .21, respectively. So, these results show that training medical students on interpersonal skills does not negate the value of selecting them on those skills via SJTs in the first place.

Finally, research has examined retest and coaching effects in an entry level medical selection context as in medical admissions there also exist a large coaching business. It has been shown that candidates who retake SJTs score on average .32 standard deviations better than one-time test-takers (Lievens, Buyse, & Sackett, 2005b). This effect size was in the same range as the one associated with cognitive tests of the exam. Similarly, experimental studies reveal that coaching has been found to raise SJT scores with at most .24 standard deviations (Cullen, Sackett, & Lievens, 2006). This value is also similar to coaching effects associated with cognitive tests (Hausknecht, Halpert, Di Paolo, & Moriarty Gerrard, 2007). A recent field study showed less optimistic results concerning the effects of coaching on SJTs.

Coaching effects have been estimated at about .5 standard deviations (Lievens, Buyse, Sackett, & Connelly, in press). However, future research is needed to ascertain whether the improvement in SJT scores is genuine or artificial, and confirming these issues in field studies is a complex research proposition.

Question 3: What Is The Incremental Validity Of Low-Fidelity And High-Fidelity Simulations Over Cognitively-Oriented Predictors In Advanced Level Selection?

Anecdotally, medics often say that '*physicians don't fail because they are not bright enough, it's usually that they have the wrong attitude*'. By contrast, selection practices in medicine have tended to focus on testing academic ability alone (Irish et al., 2011). Due to the desire to include a broader array of competencies, in some countries both low-fidelity and high-fidelity simulations have been used to measure key noncognitive competencies. Research has then examined whether these low-fidelity selection approaches indeed "deliver" what they attempt to do, namely adding incremental predictive power over and above cognitively-oriented selection predictors.

Empirical evidence from the UK in an advanced level selection context found significant positive correlations between cognitive ability tests and both the bespoke clinical problem solving test and SJT, ranging between .34 and .47. Results indicate that the cognitive ability tests on the one hand and the clinical problem solving test and SJT on the other hand measure to some extent overlapping constructs. For example, regression analyses indicated that a significant proportion of additional variance in AC performance was explained by the SJT compared to cognitive ability tests (for a comprehensive description see Koczwara et al., 2012).

As there is a public perception that junior physicians are generally above average intelligence, using methods such as cognitive ability tests for advanced level (postgraduate) selection may be rendered less applicable. However, de-selecting candidates at postgraduate

level based on cognitive ability test scores may be at odds to previous high academic success which enabled selection into medical school in the first place (see Koczwara, et al., 2012). Therefore, in advanced level medical selection, knowledge tests (instead of cognitive ability tests) are typically used for measuring an applicant's declarative knowledge (Lievens & Patterson, 2011). Here, knowledge testing can be used as a way of determining cognitive ability (Kanfer & Ackerman, 1989) since cognitive ability is thought to be a determinant of knowledge acquisition and learning (Motowidlo & Beier, 2010). In high-stakes settings, knowledge tests can assess an individual's attainment of previously learned knowledge and are developed around the assumption that the applicant has been trained on specific job-related tasks (Hunter & Hunter, 1984).

Researchers have examined whether low-fidelity simulations add validity over and above knowledge tests in advanced level selection. Patterson, Baron et al. (2009) found that in an advanced level selection context, the SJT explained an additional 11% in AC performance over a clinical knowledge test. Lievens and Patterson (2011) extended these results by showing that the SJT could also explain an additional 5.3% of the variance in job performance (instead of AC performance) over and above the clinical knowledge test. Moreover, structural equation modelling analysis specified that the clinical knowledge test (measure of declarative knowledge) had no direct effect on job performance; the relationship was fully mediated by SJT performance. The AC scores could not explain all variance in job performance that was captured by SJT, indicating that the inclusion of AC exercises did not make the use of an SJT redundant in this context. In turn, the AC could explain a unique portion of variance (2.1%) in job performance that could not be explained by the SJT or the knowledge test. This suggests that a combination of low-fidelity and high-fidelity simulations yields higher validity coefficients compared to each simulation on its own. A final yet unpublished study (Patterson, 2011) found that the SJT explained 2 to 8% additional variance

on end-of-training licensure exam performance after 3 years of training above the clinical knowledge test.

Question 4: What Is The Incremental Validity Of Low-Fidelity Simulations Over Cognitively-Oriented Predictors In Entry-Level Selection?

Similar to advanced level medical selection, the issue as to whether low-fidelity selection methods (SJTs) add predictive power over and above cognitive ability in a medical admission setting has received scant research attention. Oswald et al. (2004) demonstrated that an SJT in combination with a biodata measure had incremental validity beyond cognitive oriented college entrance tests for GPA, absenteeism and college performance ratings (up to 21.6%). Schmitt et al. (2009) extended these findings by showing that SJT in combination with a biodata measure could explain 2.9% incremental variance in four-year GPA, beyond more cognitive oriented entrance tests. Lievens et al. (2005a) confirmed these US findings in the Belgian high-stakes medical admission context. A video-based SJT could explain incremental variance over cognitive tests in GPA on interpersonal skills courses up to 7%. In their follow-up study in which they tracked students from admission to employment, Lievens and Sackett (2012) found that the video-based SJTs had incremental validity over and above the traditional cognitive tests for internship performance seven years later and job performance as a trainee physician nine years later (up to 4.6%).

In short, when we summarize the research evidence related to these four validity questions, low-fidelity simulations (SJT) seem to be valid predictors of subsequent performance in high-stakes settings. SJT is the best predictor of AC performance and end of training licensure exam as compared to knowledge tests and cognitive ability tests. In addition to knowledge tests, SJTs can explain a significant and unique portion of variance in supervisor appraisal ratings one year into training and performance at end of training

licensure exam, that cannot be captured by declarative knowledge tests and high fidelity simulations (i.e., AC).

Question 5: How Do Applicants React Towards Low-Fidelity And High-Fidelity Selection Instruments In High Stakes Settings?

In high stakes selection, where results play a critical role in gaining access to education and employment (Sackett, Schmitt, Ellingson, & Kabin, 2001), it is logical that there is a strong public interest in how physicians and specialists are selected and in defining the criteria by which we judge their competence. Hence, besides the validity of low-fidelity and high-fidelity simulations in high-stakes medical selection, a related concern is in understanding how all stakeholders and especially applicants perceive and react to the selection process. According to the meta-analysis of Hausknecht, Day, and Thomas (2007), applicant reactions are important factors related to the intention to dissuade others to take part in the selection process. Given the need to attract sufficient physicians into all specialties, this should definitely be avoided. In addition, adverse applicant reactions are associated with intentions for legal case initiation (Anderson, 2011). For instance, as there is a single career path into specialties within the UK health servce, failure at this selection gateway can severely limit career choices (Patterson & Ferguson, 2007; Patterson & Zibarras, 2011). As a result, the threat of legal case initiation is perceived to be greater than in many other occupational settings.

Koczwara et al. (2012) examined the face validity of selection instruments used for live advanced level medical selection, alongside two new cognitive ability tests, as part of a pilot study. Immediately after completing the tests, participants were asked to complete an evaluation questionnaire regarding their perceptions of the tests, based on procedural justice theory (Bauer, Truxillo, Sanchez, Craig, Ferrara & Campion, 2001). Results show the knowledge test received the most positive feedback, followed by the SJT. On the other hand,

the cognitive ability tests were not positively received as they were perceived to have exceptionally low job relevance and to offer little opportunity to demonstrate candidates' ability. Similar findings have emerged in Belgium in an entry level admission procedure (Lievens, Coetsier, & Buyse, 2001). The video-based SJT was perceived as significantly more face valid than the cognitive ability tests.

In a comprehensive study, Patterson, Zibarras, Carr, Irish, and Gregory (2011) investigated applicant reactions towards the whole advanced level medical selection procedure for both the shortlisting and the assessment center selection stage. Data were collected from candidates during three annual recruitment rounds (2007-2009, N=approx. 6,000 candidates per annum) of general practitioners in the UK. Following their participation in each stage of the selection process, candidates were invited to complete an evaluation questionnaire. All selection methods used in the selection process were considered to be job relevant and both selection stages were considered fair in terms of their formal test characteristics and the interpersonal treatment. However, as would be expected, the high fidelity assessment center phase received more positive ratings than the shortlisting stage.

In relation to the shortlisting stage, further analyses indicated that the low fidelity clinical knowledge test was considered to be highly job relevant over three consecutive years, whilst perceptions of job relevance of the SJT were consistently lower, although this improved marginally over the three years. Perceptions of the overall fairness of the shortlisting process also improved in 2009 compared to 2008. In relation to the assessment center, findings suggested that job relevance perceptions of the high fidelity selection methods improved significantly between 2007 and subsequent years. In each year however, when comparing the three assessment center exercises, the simulated patient consultation was perceived to be the most job relevant by candidates. Perceptions of fairness of the assessment center stage overall were comparatively high in each year, with no differences in candidate

perceptions between years. Overall, these findings show that the advanced level selection process was received positively by candidates. All selection methods were considered job relevant and each stage of the process was considered fair. So, these results support previous research suggesting higher-fidelity selection methods are rated more positively than lower-fidelity methods (e.g. Steiner & Gilliland, 1996). Findings also imply what has been labeled a "justice dilemma" (Cropanzano & Konovsky, 1995) because the selection method from the shortlisting stage that had the highest predictive validity for job performance – the SJT – had lower face validity than the knowledge test. The results indicate that the junior physicians perceive the knowledge/cognitively-oriented test to be more relevant to their role compared to the SJT that focused on important non-cognitive attributes. In practice, by seeking feedback from candidates, recruiters were alerted to this potential dilemma and specific interventions were introduced. The aim was to increase information given to candidates about the SJT regarding its relevance to the general practitioner role, before, during and after the shortlisting stage. Examples of these interventions included providing detailed information via a national recruitment website about the reasons for using an SJT, and what the test is assessing (i.e. attributes beyond clinical knowledge). A separate section of the website also provides a summary of the research evidence on the use of SJTs in selection. These communication interventions had a positive effect as candidate perceptions of the SJT improved year on year.

Discussion

This paper aimed to review recent research about the effectiveness of measuring noncognitive predictors via low-fidelity (SJT) and/or high-fidelity (AC) selection methods. In light of the high-stakes context of medical selection practices we scrutinized the low and high fidelity selection methods in terms of predictive validity, incremental validity, and candidate perceptions. Hereby we focused on published and unpublished medical selection research in

primarily two countries, namely the UK (advanced level medical selection) and Belgium (entry-level medical selection).

One general conclusion from this review is that the cognitive/knowledge tests and the SJT (focusing on non-cognitive attributes) are all significant predictors of key criteria (i.e., relevant performance in the short, medium, and long term). Logically, the best prediction of performance outcomes is a combination of all the methods. As another important conclusion of the review, the low fidelity SJT has incremental predictive power over cognitively oriented tests. Furthermore, high fidelity AC exercises add incremental validity over the low fidelity (and less costly) selection methods. When exploring the predictive validity of selection methods and candidate reactions, bespoke selection tests such as the SJT and a declarative knowledge test (clinical problem solving test) are marginally better predictors of performance than the assessment centre exercises. These two bespoke tests are also considered by candidates to have significantly higher face validity than two cognitive ability tests. Similar applicant perception results are found in the medical admissions context. These findings have important implications for both practice and theory in high stakes selection.

Implications for System Design, Policy, and Practice in High Stakes Selection

The research exploring the selection systems reported in this paper represent long-term research programs to create selection systems that are valid, fair, and transparent. In practice, the selection methods have been developed incrementally over several years. Findings also suggest that the bespoke selection tests have strong predictive validity that can translate into significant gains in utility. Although the focus of this paper has centred on the validity of various selection methods, we argue that the design and delivery of this research represents a major organizational change program, conducted in partnership with key stakeholders in the medical profession (see Plint & Patterson, 2010). Stakeholder involvement has to be significant at each step of the design process, from the job analysis, to the design of the selection methods, through

to the design of the outcome variables - and this stakeholder involvement constitutes an important factor in the successful implementation in a medical context that historically has paid little attention to selection research (Irish et al., 2011). We argue in favour of the importance of a multi-source, multi-method bespoke job analysis study, as the cornerstone to delivering a valid and credible selection system.

The research reviewed here demonstrates the need to explore validity of a selection system from various perspectives in high stakes settings. For example, we know that cognitive ability tests tend to be the best single predictor of future job performance across many occupational groups internationally (Schmidt & Hunter, 1998; Salgado, et al, 2003a; Salgado, et al, 2003b). However, there is also a need to conduct validation studies from both the organization and the candidates' perspectives so that an organization can assess the extent to which selection methods are positively received (Bauer et al., 2001; Patterson et al., 2011; Truxillo, Bauer, Campion, & Paronto, 2002). This issue is especially important in high stakes settings for advanced level job roles. We reviewed studies that showed that cognitive ability tests were significantly less positively received by candidates than bespoke selection methods. Reactions towards the two cognitive ability tests were so negative that stakeholder reactions were considered important in this context. De-selecting a junior physician on the basis of a cognitive ability test at an advanced level was not seen as acceptable and legitimate. This conforms to previous applicant reaction research in other occupations where candidates perceive cognitive ability tests as irrelevant to the job role (Anderson & Witvliet, 2008; Nikolaou & Judge, 2007; Smither, Reilly, Millsap, Pearlman, & Stoffey, 1993). Candidates' negative perceptions can lead to undesirable outcomes such as losing competent candidates from the selection process (Schmit & Ryan, 1997), which can ultimately reduce the utility of the process (Murphy, 1986).

In overview, the results presented here lead the authors to acknowledge the importance of wider stakeholder acceptance, beyond the organization and candidate. When designing selection systems in high stakes settings we promote the concept of *'political validity'* (Patterson & Zibarras, 2011) which relates to the fact that many stakeholder groups, beyond the organization itself, may influence the design and development of a selection system. For example, in medical selection across the globe, there are a variety of perspectives that must be taken into account including, among others: the professional trade union; the regulators; specific professional bodies, such as the Medical Royal Colleges; employer associations, the general public; politicians and the government. Therefore, the design, development and delivery of any selection system in medicine requires buy-in and acceptance from a huge variety of important stakeholders (see Prideaux et al., 2011), often with competing views about how selection practices should be conducted. Some current selection practices in medicine have little demonstrable evidence of predictive validity (e.g., a lottery system for entry into medical school, or referees reports for advanced level selection) but these practices may be deemed credible and appropriate by a range of important stakeholders. For example, a lottery system for entry into medical school fulfils a political agenda of widening participation, since variables such as socio-economic status; parental earnings and access to education are removed. Similarly, unstructured referees reports for entry into advanced level medicine fulfils the political agenda for employers as senior medics personal judgement of a trainee's character is perceived as more credible than results from say a 'one-off' simulated patient consultation in an assessment center exercise. Balancing the predictive validity with political validity in high-stakes selection is likely to increase the likelihood of acceptance of effective evidence-based selection practices.

Implications for Theory Underlying High-fidelity and Low-fidelity Simulations

Results from this review help integrate research exploring the validity of knowledge tests, high-fidelity, and low-fidelity simulations in advanced level selection by investigating

how each of these predictors work in combination to predict job performance in the long-term. Specifically, the knowledge determinants of the low-fidelity SJT and the high-fidelity assessment center simulations differ. Our review showed that procedural knowledge as measured by an SJT fully mediated the effects of declarative knowledge (measured by the clinical problem solving test) on job performance. Conversely, the high fidelity assessment center exercises are not related to declarative knowledge. However, they remain valid and essential for predicting important non-cognitive job performance domains.

The results presented in this review also contribute to recent theoretical developments regarding the construct-related validity of low-fidelity simulations in selection, and in particular SJTs. Motowidlo and Beier (2010) suggest that the procedural knowledge measured in SJTs is made up of implicit trait policies (ITPs) and specific job knowledge. It is likely that personality traits interact with both general and specific experiences and these experiences lead to procedural knowledge about what is effective behaviour captured by an SJT. ITPs are described as beliefs about the costs/benefits of expressing certain traits (for example, knowing that generally being more agreeable is likely to be better across a range of situations). ITPs are shaped by experiences in fundamental socialisation processes (for instance, parental modelling) that teach the utility of agreeable expressions, such as helping others in need or disagreeable actions, such as showing selfish preoccupation with one's own interests, advancing one's own interests at others' expense. In this context, we propose that the ITPs measuring SJTs in the advanced level selection context are shaped by experiences within the medical education and training pathway, where medical students learn about the behaviours associated with being a "*competent physician*". Conversely, the validity of the SJT in entry-level selection can be understood from the fact that this type of SJT captures primarily ITPs shaped by experiences in fundamental socialisation processes instead of by medical training and experience.

Limitation and Avenues For Future Research

It should be noted that there are some limitations to the research reviewed in this paper, which pertain primarily to selection methods developed in the UK and Belgium. Strictly speaking, the implications drawn in terms of theory and practice are therefore relevant only to this context. However, the selection methods studied are fairly common among high-stakes (and other) selection process and so to the extent that other selection methods are similar, these results are likely to be generalizable.

We propose a number of recommendations for future research. First, from a theoretical perspective, there is much to be gained from investigating the link between ITPs and low-fidelity simulations, high-fidelity simulations, and performance in medical school, and later job performance. This dovetails exploring research questions such as: how do ITPs develop over the training pathway for physicians? Do ITPs facilitate domain specific knowledge acquisition and how might this operate when exploring different medical specialties such as general practice versus surgery? Can individuals be equally successful at the same job whilst holding different ITPs? Second, it is important to examine the subgroup differences associated with low-fidelity versus high-fidelity simulation instruments in a high-stakes context. In many countries, the medical school admissions process has been criticized for failing to address a widening participation agenda (e.g., Barr, 2010; Schwartz, 2004) since evidence consistently shows that medical students come from higher socioeconomic classes (Garlick & Brown, 2008; Ip & McManus, 2008).

Finally, research in Canada found that a high-fidelity AC administered when entering college, including a number of role playing tests and structured situational interviews, is predictive for a variety of performance measures (e.g., Eva et al., 2009). It is not clear whether an AC can explain incremental variance beyond an SJT for college admission. Moreover, given recent findings on effects of coaching on SJTs, it would be interesting to

investigate whether people who took SJT coaching also score differently on the AC, in order to explore whether the improvement in SJT scores is genuine or artificial.

Conclusion

Medicine is a high stakes profession that historically has been relatively under-researched when exploring selection system design and the validity of various selection methods. This paper reviewed long-term research programmes in Europe to document recent innovations in medical selection practices. For a profession that has tended to focus on academic/cognitive abilities in selection, we demonstrate the added value of testing non-cognitive attributes such as interpersonal skills, empathy, integrity and teamwork via both low-fidelity (SJT) and high-fidelity (AC) selection methods in order to predict successful job performance. In designing high stakes selection, there is also a clear need to consider not just the predictive validity of selection methods but also to acknowledge the socio-political context within which selection systems are designed and implemented.

References

Anderson, N. (2011). Perceived Job Discrimination: Toward a model of applicant propensity to case initiation in selection, *International Journal of Selection and Assessment*, 19, 229- 244.

Anderson, N., Salgado, J. F., & Hulsheger, U. R. (2010). Applicant reactions in selection: Comprehensive meta-analysis into reaction generalization versus situational specificity. *International Journal of Selection and Assessment,* 18, 291-304.

Anderson, N., & Witvliet, C. (2008). Fairness reactions to personnel selection methods: An international comparison between the Netherlands, the United States, France, Spain, Portugal, and Singapore. *International Journal of Selection and Assessment*, *16*, 1-13.

Arthur, W., Day, E. A., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology, 56*, 125-154.

Barr, D. A. (2010). *Questioning the premedical paradigm; enhancing diversity in the medical profession a century after the Flexner report*. Baltimore, MA: John Hopkins University Press.

Bauer, T. N., Truxillo, D. M., Sanchez, R. J., Craig, J. M., Ferrara, P., & Campion, M A. (2001). Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel Psychology*, *54*, 387-419.

Christian, M., Edwards, B., & Bradley, J. (2010). Situational judgement tests: constructs assessed and a meta-analysis of their criterion-related validities. *Personnel Psychology*, *63*, 83-117.

Cropanzano, R., & Konovsky, M. A. (1995). Resolving the justice dilemma by improving the outcomes: The case of employee drug screening. *Journal of Business and Psychology*, *10*, 221-243.

Cullen, M. J., Sackett, P. R., & Lievens, F. (2006). Threats to the operational use of situational judgment tests in the college admission process. *International Journal of Selection and Assessment, 14*, 142-155.

Eva, K. W., Reiter, H. I., Trinh, K., Wasi, P., Rosenfeld, J., & Norman, G. R. (2009). Predictive validity of the multiple mini-interview for selecting medical trainees. *Medical Education, 43*, 767-775.

Garlick, P., & Brown, G. (2008). Widening participation in medicine. *British Medical Journal*, *336*, 1111.

Greenhalgh, T. (2010). Widening participation: Say no to nepotism. *British Medical Journal*, *341*, c6130.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639-683.

Hausknecht, J.P., Halpert, J.A., Di Paolo, N.T., & Moriarty Gerrard, M.O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology, 92*, 373-385.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*, 72-98.

Ip, H., & McManus, I. (2008). Replies to increasing diversity among clinicians. *British Medical Journal*, *336*, 1082.

Irish, B., Carr, A., Sowden, D., Douglas, N., & Patterson, F. (2011). Recruitment into specialty training in the UK. *British Medical Journal*, *342, 19*.

Jefferis, T. (2007). Selection for specialist training: What can we learn from other countries? *British Medical Journal, 334*, 1302–1304.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, *74*, 657-690.

Klein, C., DeRouin, R. E., & Salas, E. (2006). Uncovering workplace interpersonal skills: A review, framework, and research agenda. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology* (Vol. 21, pp. 80-126). New York: Wiley & Sons, Ltd.

Koczwara, A., Patterson, F., Zibarras, L., Kerrin, M., Irish, B., & Wilkinson, M. (2012). Evaluating cognitive ability, knowledge tests and situational judgement tests for postgraduate selection. *Medical Education, 46*, 399-408.

Lievens, F., Buyse, T., & Sackett, P. R. (2005a). The operational validity of a video-based situational judgment test for medical college admissions: Illustrating the importance of matching predictor and criterion construct domains. *Journal of Applied Psychology, 90*, 442-452.

Lievens, F., Buyse, T., & Sackett, P. R. (2005b). Retest effects in operational selection settings: Development and test of a framework. *Personnel Psychology, 58*, 981-1007.

Lievens, F., Buyse, T., Sackett, P. R., & Connelly, B. S. (in press). The effects of coaching on situational judgment tests in high-stakes selection. *International Journal of Selection and Assessment.*

Lievens, F., Coetsier, P., & Buyse, T. (2001). Het toelatingsexamen 'Arts en Tandarts' in Vlaanderen: voorspellende validiteit en reacties van de studenten. *Tijdschrift voor Medisch Onderwijs*, *5*, 186-195.

Lievens, F., & Patterson, F. (2011). The validity and incremental validity of knowledge tests, low-fidelity simulations, and high-fidelity simulations for predicting job performance in advanced-level high-stakes selection. *Journal of Applied Psychology, 96*, 927-940.

Lievens, F., Peeters, H., & Schollaert, E. (2008). Situational judgment tests: a review of recent research. *Personnel Review, 37*, 426-441.

Lievens, F., & Sackett, P. R. (2012). The validity of interpersonal skills assessment via situational judgment tests for predicting academic success and job performance. *Journal of Applied Psychology, 97*, 460-468.

McDaniel, M., Morgeson, F. P., Finnegan, E. B., Campion, M. A., & Braverman, E. P. (2001). Use of Situational Judgment Tests to Predict Job Performance: A Clarification of the Literature. *Journal of Applied Psychology*, *86*, 730-740.

McDaniel, M., Hartman, N., Whetzel, D., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology*, *60*, 63-91.

McManus, I. C., Smithers, E., Partridge, P., Keeling, A., & Fleming, P. (2003). A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. *BMJ,* 327(7407), 139–142.

McManus, I. C., Powis, D., Wakeford, R., Ferguson, E., James, D., & Richards, P. (2005). Intellectual aptitude tests and A levels for selecting UK school leaver entrants for medical school. *BMJ,* 331(7516), 555–559.

Motowidlo, S. J., & Beier, M. E. (2010). Differentiating specific job knowledge from implicit trait policies in procedural knowledge measured by a situational judgment test. *Journal of Applied Psychology*, *95*, 321-333.

Motowidlo, S. J., Dunnette, M. D., & Carter, G. W. (1990). An alternative selection procedure: The low-fidelity simulation. *Journal of Applied Psychology*, *75*, 640-647.

Murphy, K. R. (1986). When your top choice turns you down: Effect of rejected offers on the utility of selection tests. *Psychological Bulletin*, *99*, 133-138.

Nikolaou, I., & Judge, T. A. (2007). Fairness reactions to personnel selection techniques in Greece: The role of core self-evaluations. *International Journal of Selection and Assessment*, *15*, 206-219.

Oswald, F. L., Schmitt, N., Kim, B. H., Ramsay, L. J., & Gillespie, M. A. (2004). Developing a biodata measure and situational judgment inventory as predictors of college student performance. *Journal of Applied Psychology*, *89*, 187-207.

Patterson, F. (2011). Longitudinal validation of low fidelity tests for high stakes selection: Implications for theory, policy & practice. *EAWOP Small Group Meeting,* Athens, June.

Patterson, F., Baron, H., Carr, V., Plint, S., & Lane, P. (2009). Evaluation of three short-listing methodologies for selection into postgraduate training in general practice. *Medical Education*, *43*, 50-57.

Patterson, F., Carr, V., Zibarras, L., Burr, B., Berkin, L., Plint, S., Irish, B., Gregory, S. (2009). New machine-marked tests for selection into core medical training: evidence from two validation studies. *Clinical Medicine*, *9*, 417-420.

Patterson, F., & Ferguson, E. (2007). *Selection into Medical Education and Training*. Edinburgh: Association for the Study of Medical Education.

Patterson, F., Ferguson, E., Norfolk, T., & Lane, P. (2005). A new selection system to recruit general practice registrars: preliminary findings from a validation study. *British Medical Journal*, *330*, 711-714.

Patterson, F., & Zibarras, L. (2011). Exploring the construct of perceived job discrimination in selection. *International Journal of Selection and Assessment*, *19*, 259-265.

Patterson, F., Zibarras, L., Carr, V., Irish, B., & Gregory, S. (2011). Evaluating candidate reactions to selection practices using organisational justice theory. *Medical Education, 45*, 289-297.

Plint, S., & Patterson, F. (2010). Identifying critical success factors for designing selection processes into postgraduate specialty training: the case of UK general practice. *Postgraduate Medical Journal*, *86*, 323-327.

Prideaux, D., Roberts, C., Eva, K., Centeno, A., McCrorie, P., McManus, C., Patterson, F., Powis, D., Tekian, A. (2011). Assessment for selection for the health care professions and specialty training: Consensus statement and recommendations from the Ottawa 2010 Conference. *Medical Teacher, 33*, 215-223.

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology, 59*, 419-445.

Sackett, P., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, *56*, 302-318.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003a). International validity generalization of GMA and cognitive abilities: A European community meta- analysis. *Personnel Psychology, 56*, 573-605.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., De Fruyt, F., & Rolland, J. P. (2003b). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology, 88*, 1068-1081.

Schmidt, F., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262-274.

Schmit, M. J., & Ryan, A. M. (1997). Applicant withdrawal: the role of test-taking attitudes and racial differences. *Personnel Psychology*, *50*, 855-876.

Schmitt, N., Billington, A., Keeney, J., Oswald, F., Pleskac, T., Sinha, R., & Zorzie, M. (2009). Prediction of four-year college student performance using cognitive and noncognitive predictors and the impact on demographic status of admitted students. *Journal of Applied Psychology*, *94*, 1479–1498.

Schwartz, S. (2004). *Fair admissions to higher education: recommendations for good practice.* Nottingham: Department for Education and Skills.

Smither, J. W., Reilly, R. R., Millsap, R. E., Pearlman, K., & Stoffey, R. W. (1993). Applicant reactions to selection procedures. *Personnel Psychology*, *46*, 49-76.

Steiner, D. D., & Gilliland, S. W. (1996). Fairness reactions to personnel selection techniques in France and the United States. *Journal of Applied Psychology*, *81*, 134-141.

Truxillo, D. M., Bauer, T. N., Campion, M. A., & Paronto, M. E. (2002). Selection fairness information and applicant reactions: A longitudinal field study. *Journal of Applied Psychology*, *87*, 1020-1031.

Whetzel, D. L., & McDaniel, M. A. (2009). Situational judgment tests: An overview of current research. *Human Resource Management Review, 19*, 188-202.

Figure 1: The three-stage advanced level selection methodology for medicine

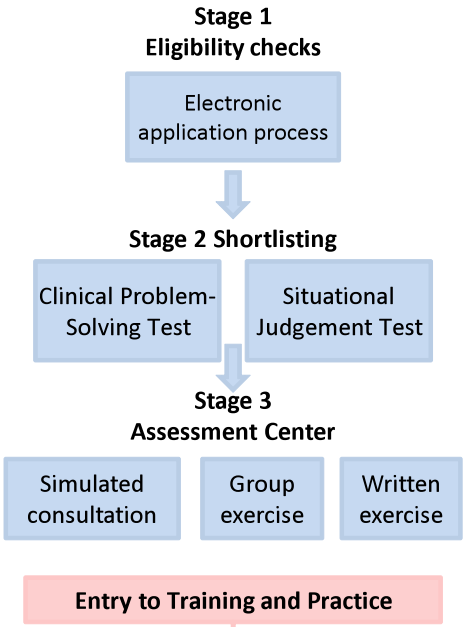

Table 1.

Example Items for the Clinical Problem-Solving Test and the Situational Judgement Test

Example Clinical Problem-solving item	**Example SJT item**
Reduced Vision A. Basilar migraine B. Cerebral tumor C. Cranial arteritis D. Macular degeneration E. Central retinal artery occlusion F. Central retinal vein occlusion G. Optic neuritis (demyelinating) H. Retinal detachment I. Tobacco optic neuropathy For each patient below select the SINGLE most likely diagnosis from the list above. Each option may be selected once, more than once or not at all. 1. A 75 year old man, who is a heavy smoker, with a blood pressure of 170/105, complains of floaters in the left eye for many months and flashing lights in bright sunlight. He has now noticed a "curtain" across his vision.	You are reviewing a routine drug chart for a patient with rheumatoid arthritis during an overnight shift. You notice that your consultant has inappropriately prescribed methotrexate 7.5mg daily instead of weekly. *Rank in order the following actions in response to this situation (1= Most appropriate; 5= Least appropriate).* A. Ask the nurses if the consultant has made any other drug errors recently B. Correct the prescription to 7.5mg weekly C. Leave the prescription unchanged until the consultant ward round the following morning D. Phone the consultant at home to ask about changing the prescription E. Inform the patient of the error

Table 2.

Descriptive Statistics and Correlations (Patterson 2011)

N=2292	***Mean***	***SD***	**1.**	**2.**	**3.**	**4.**
Predictors						
1. Clinical problem solving test	80.08	8.14				
2. Situational judgement test	640.13	31.66	.40			
3. Assessment center	3.34	0.36	.24	.32		
Criteria						
4. End of training applied knowledge test	0.26	0.90	.73 *(.85)*	.43 *(.69)*	.24 *(.41)*	
5. End of training clinical skills assessment	0.20	0.80	.38 *(.55)*	.43 *(.57)*	.32 *(.41)*	.41

Note. Correlations between parentheses were corrected for multivariate range restriction. All correlations are significant at $p < .001$. Criteria measures are the two-part licensing examination for Membership of the Royal College of General Practice (MRCGP).